



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08		A1	(11) International Publication Number: WO 98/07864 (43) International Publication Date: 26 February 1998 (26.02.98)
<p>(21) International Application Number: PCT/GB97/02273</p> <p>(22) International Filing Date: 22 August 1997 (22.08.97)</p> <p>(30) Priority Data: 9617671.4 23 August 1996 (23.08.96) GB 9625996.5 13 December 1996 (13.12.96) GB </p> <p>(60) Parent Application or Grant</p> <p>(63) Related by Continuation US 08/782,893 (CJP) Filed on 27 December 1996 (27.12.96) </p> <p>(71) Applicants (for all designated States except US): MICROBIOLOGICAL RESEARCH AUTHORITY CAMR (CENTRE FOR APPLIED MICROBIOLOGY & RESEARCH) (GB/GB); Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED (GB/GB); 14 Kensington Square, London W8 5HH (GB).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles (GB/GB); Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUITINN,</p>		Conrad, Padraig [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). <p>(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
<p>(54) Title: RECOMBINANT TOXIN FRAGMENTS</p> <p>(57) Abstract</p> <p>A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BR	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BW	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KR	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Republic of Korea	RO	Romania		
CU	Cuba	LC	Kazakhstan	RU	Russian Federation		
CZ	Czech Republic	LI	Saint Lucia	SD	Sudan		
DE	Germany	LK	Liechtenstein	SE	Sweden		
DK	Denmark	LR	Sri Lanka	SG	Singapore		
EE	Estonia		Liberia				

RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the H_C domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the H_N domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The H_N domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the H_C, amino acid residues 449-871 for the H_N and residues 1-448 for the LC. Digestion with trypsin effectively degrades the H_C domain of the BoNT/A to generate a non-toxic fragment designated LH_N.

which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

(A) clostridial neurotoxin light chain:

-a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.

(B) clostridial neurotoxin heavy chain H_N domain:

-a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.

-the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.

-the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.

-the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.

(C) clostridial neurotoxin heavy chain H_C domain.

-a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of cl stridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H_C domain.

For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a H_N domain or to the functions of a H_N domain are references to this property or properties. The second domain is not required to exhibit other properties of the H_N domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin - this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from *Aplysia* (sea slug), *Drosophila* (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain H_N portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the H_N domain. Teachings of regions within the H_N responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment, though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a H_N domain or fragment, (ii) a C-terminal extension to a H_N domain or fragment, (iii) a modification to a H_N domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an *in vitro* assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated H_C of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated H_C of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endopeptidase activity of the composition will, of course, also be determined by the amount of the derivative that is present.

While it is known to treat naturally produced clostridial toxin to remove the H_c domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH₄₂₃/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH₇₂₈/B (SEQ ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host:

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The LH_{N/A} derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH₄₂₃/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_C domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a '/' to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nomenclature,

- 13 -

LH_{423}/A = SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;

${}_2LH_{423}/A$ = a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

${}_2L_{12}H_{423}/A$ = a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain;

${}_2L_{FXa/2}H_{423}/A$ = a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

${}_2L_{FXa/2}H_{423}/A\text{-IGF-1}$ = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH_{423}/A ;

Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designated ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇,Y) in an *in vitro* peptide cleavage assay;

Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;

Fig. 5 shows the transition region of an embodiment of the invention designated L₄H₄₂₃/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the Eco 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;

Fig. 6 shows the transition region of an embodiment of the invention designated L_{FXa3}H₄₂₃/A illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H_N sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;

Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated L_{FXa3}H₄₂₃/A-IGF-1, a fusion protein; the IGF-1 sequence begins at position G₈₈₂;

Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated L_{FXa3}H₄₂₃/A-CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Q₈₈₂;

Fig. 9 shows the C-terminal portion of the amino acid sequence of an

- 15 -

embodiment of the invention designated $L_{Fx\alpha 3}H_{423}/A-ZZ$, a fusion protein; the C-terminal ZZ sequence begins at position A_{890} immediately after a gene-nase recognition site (underlined);

show schematic representations of manipulations of
Figs. 10 & 11 polypeptides of the invention; Fig. 10 shows LH_{423}/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12 shows the trypsin sensitive activation region of a polypeptide of the invention;

Fig. 13 shows Western blot analysis of recombinant LH_{107}/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH_{107}/B ; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH_{107}/B .

The sequence listing that accompanies this application contains the following sequences:-

<u>SEQ ID NO:</u>	<u>Sequence</u>
1	DNA coding for LH_{423}/A

- 16 -

2 LH₄₂₃/A

3 DNA coding for ₂₃LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y), of which an N-terminal portion is shown in Fig. 4.

4 ₂₃LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y)

5 DNA coding for ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y), of which an N-terminal portion is shown in Fig.4

6 ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y)

7 DNA coding for native BoNT/A according to Binz et al
native BoNT/A according to Binz et al

9 DNA coding for L₁₄H₄₂₃/A

10 L₁₄H₄₂₃/A

11 DNA coding for L_{FX₀}/₃H₄₂₃/A

12 L_{FX₀}/₃H₄₂₃/A

13 DNA coding for L_{FX₀}/₃H₄₂₃/A-IGF-1

14 L_{FX₀}/₃H₄₂₃/A-IGF-1

15 DNA coding for L_{FX₀}/₃H₄₂₃/A-CtxA14

16 L_{FX₀}/₃H₄₂₃/A-CtxA14

17 DNA coding for L_{FX₀}/₃H₄₂₃/A-ZZ

18 L_{FX₀}/₃H₄₂₃/A-ZZ

19 DNA coding for LH₇₂₈/B

20 LH₇₂₈/B

21 DNA coding for LH₄₁₇/B

22 LH₄₁₇/B

23 DNA coding for LH₁₀₇/B

24 LH₁₀₇/B

25 DNA coding for LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y)

26 LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y)

27 DNA coding for LH₄₁₇/B wherein the first 274 bases are

28

modified to have an *E.coli* codon bias

DNA coding for LH₄₂₃/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_C) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an *E. coli* codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique *Kpn*I restriction site. The remainder of the LH₄₂₃/A coding sequence was PCR amplified from total chromosomal DNA from *Clostridium botulinum* and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZ218, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, *E. coli* harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced by addition of isopropylthio- β -D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH₄₂₃/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST- LH₄₂₃/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH₄₂₃/A is recovered in the non-adsorbed material.

Construct variants

A variant of the molecule, LH₄₂₃/A (Q₂E,N₂₆K,A₂₇,Y) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH₄₂₃/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A .

Two further variants of the gene sequence that have been expressed and the corresponding products purified are ₂₃LH₄₂₃/A (Q₂E,N₂₆K,A₂₇,Y) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇,Y) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a Eco 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the Eco 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, L₁₄H₄₂₃/A (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of LH₄₂₃/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L_{FXa3}H₄₂₃/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L₁₄H₄₂₃/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated; and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of L_{FXa3}H₄₂₃/A have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1) L_{FXa3}H₄₂₃/A-IGF-1 (SEQ ID NO: 14) , in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2) L_{FXa3}H₄₂₃/A-CtxA14 (SEQ ID NO: 16) , in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) L_{FXa3}H₄₂₃/A-ZZ (SEQ ID NO: 18) , in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptid

incorporating a domain equivalent to the light chain of botulinum n urotoxin type A and a domain equivalent to the H_N domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of ₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST-₂LH₄₂₃/A (Q₂E,N₂₆K,A₂₇Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 µg/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 µg/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

₂LH₄₂₃/A is more stable in the presence of trypsin and more active in the *in vitro*

peptide cleavage assay than is LH_{423}/A . Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH_{728}/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide, LH_{417}/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native LH_{N/A}.

A gene has also been assembled coding for a variant polypeptide, LH_{107/B} (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-₂LH₄₂₃/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA 11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 µg of GST-₂LH₄₂₃/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH_N/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na₂HPO₄ anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 µg (5X10⁻⁶ g) to 5 picograms (5X10⁻¹² g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- GST-₂LH₄₂₃/A antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-₂LH₄₂₃/A antiserum.

Survivors On Day	<u>Botulinum Toxin/mouse</u>							Control (no toxin)
	0.5 μ g	0.005 μ g	0.0005 μ g	0.5ng	0.005ng	5pg		
1	0	4	4	4	4	4	4	4
2	-	4	4	4	4	4	4	4
3	-	4	4	4	4	4	4	4
4	-	4	4	4	4	4	4	4

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

Survivors On Day	<u>Botulinum Toxin/mouse</u>							Control (no toxin)
	0.5 μ g	0.005 μ g	0.0005 μ g	0.5ng	0.005ng	5pg		
1	0	0	0	0	0	0	2	4
2	-	-	-	-	-	-	0	4
3	-	-	-	-	-	-	-	4
4	-	-	-	-	-	-	-	4

Example 4

Expression of recombinant LH₁₀₇/B in *E. coli*.

As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide LH₁₀₇/B (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in *E. coli* BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for LH₄₂₃/A.

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
(B) STREET: Centre For Applied Microbiology And Research,
Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(A) NAME: THE SPEYWOOD LABORATORY LIMITED
(B) STREET: 14 Kensington Square
(C) CITY: London
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): W8 5HH

(A) NAME: FOSTER; Keith Alan
(B) STREET: Centre For Applied Microbiology And Research,
Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(A) NAME: QUINN; Conradi Padraig
(B) STREET: Centre For Applied Microbiology And Research,
Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(A) NAME: SHONE; Clifford Charles
(B) STREET: Centre For Applied Microbiology And Research,
Porton Down
(C) CITY: Salisbury
(D) STATE: Wiltshire
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): SP4 0JG

(ii) TITLE OF INVENTION: Recombinant Toxin Fragments

(iii) NUMBER OF SEQUENCES: 28

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2616 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT.	48
Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
1 5 10 15	
GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG	96
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
20 25 30	
GTC AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC	144
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
35 40 45	
GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA	192
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	
50 55 60	
GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA	240
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	
65 70 75 80	
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG	288
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	
85 90 95	
CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC	336
Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	
100 105 110	
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG	384
Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	
115 120 125	
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC	432
Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr	
130 135 140	
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT	480
Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile	
145 150 155 160	
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG	528
Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr	
165 170 175	
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC	576
Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	
180 185 190	
ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG	624
Thr Phe Gly Phe Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu	
195 200 205	
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG	672
Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu	
210 215 220	
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC	720
Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn	
225 230 235 240	

CGC GTG TTC AAG AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTC ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 455 460	1392
AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480	1440
ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT AGT TTA Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	1488
GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510	1536

GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560	1680
CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640	1920
TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT GTA GGT GCT TTA Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655	1968
ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720	2160
GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735	2208
GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750	2256
CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT TTT AAT ATT GAT GAT Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765	2304
TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780	2352

- 33 -

AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT TCT ATG Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800	2400
ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT CTT AAA Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815	2448
GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA ATT GGT Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830	2496
CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT ACA GAT Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845	2544
ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA TTA TCT Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860	2592
ACA TTT ACT GAA TAT ATT AAG TAA Thr Phe Thr Glu Tyr Ile Lys * 865 870	2616

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 872 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95
Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110
Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125
Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

- 34 -

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190
 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205
 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220
 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240
 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255
 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
 450 455 460
 Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480
 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525

- 35 -

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750

Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
 805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
 850 855 860

Thr Phe Thr Glu Tyr Ile Lys *
 865 870

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2685 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2685

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGA TCC CCA GGA ATT CAT ATG ACG TCG ACG CGT CTG CAG AAG CTT CTA Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu 1 5 10 15	48
GAA TTC GAG CTC CCG GGT ACC ATG GAG TTC GTG AAC AAG CAG TTC AAC Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30	96
TAT AAG GAC CCT GTA AAC GGT GTT GAC ATT GCC TAC ATC AAA ATT CCA Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro 35 40 45	144
AAG TAC GGC CAG ATG CAG CCG GTG AAG GCT TTC AAG ATT CAT AAC AAA Lys Tyr Gly Gln Met Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60	192
ATC TGG GTT ATT CCG GAA CGC GAT ACA TTT ACG AAC CCG GAA GAA GGA Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80	240
GAC TTG AAC CCG CCG CCG GAA GCA AAG CAG GTG CCA GTT TCA TAC TAC Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95	288
GAT TCA ACC TAT CTG AGC ACA GAC AAC GAG AAG GAT AAC TAC CTG AAG Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys 100 105 110	336
GGA GTG ACC AAA TTA TTC GAG CGT ATT TAT TCC ACT GAC CTG GGC CGT Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg 115 120 125	384
ATG CTG CTG ACC TCA ATC GTC CGC GGA ATC CCA TTT TGG GGT GGC AGT Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser 130 135 140	432
ACC ATT GAC ACG GAG TTG AAG GTT ATT GAC ACT AAC TGC ATT AAC GTG Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val 145 150 155 160	480
ATC CAA CCA GAC GGT AGC TAC AGA TCT GAA GAA CTT AAC CTC GTA ATC Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile 165 170 175	528
ATC GGG CCC TCC GCG GAC ATT ATC CAG TTT GAG TGC AAG AGC TTT GGC Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly 180 185 190	576
CAC GAA GTG TTG AAC CTG ACG CGT AAC GGT TAC GGC TCT ACT CAG TAC His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr 195 200 205	624

ATT CGT TTC AGC CCA GAC TTC ACG TTC GGT TTC GAG GAG AGC CTG GAG Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu 210 215 220	672
GTT GAT ACC AAC CCG CTG TTG GGT GCA GGC AAG TTC GCA ACT GAT CCA Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro 225 230 235 240	720
GCG GTG ACC CTG GCA CAC GAG CTG ATC CAC GCC GGT CAT CGT CTG TAT Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr 245 250 255	768
GCG ATT GCG ATT AAC CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 260 265 270	816
TAC TAC GAG ATG AGT GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACG Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr 275 280 285	864
TTC GGT GGC CAT GAT GCG AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu 290 295 300	912
TTC CGT CTG TAC TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu 305 310 315 320	960
AAC AAG GCT AAG TCC ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met 325 330 335	1008
AAA AAT GTT TTT AAA GAG AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly 340 345 350	1056
AAA TTT TCG GTA GAT AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu 355 360 365	1104
ACA GAG ATT TAC ACA GAG GAT AAT TTT GTT AAG TTT TTT AAA GTA CTT Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu 370 375 380	1152
AAC AGA AAA ACA TAT TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn 385 390 395 400	1200
ATA GTA CCT AAG GTA AAT TAC ACA ATA TAT GAT GGA TTT AAT TTA AGA Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg 405 410 415	1248
AAT ACA AAT TTA GCA GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn 420 425 430	1296
AAT ATG AAT TTT ACT AAA CTA AAA AAT TTT ACT GGA TTG TTT GAA TTT Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe 435 440 445	1344
TAT AAG TTG CTA TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser 450 455 460	1392
TTA GAT AAA GGA TAC AAT AAG GCA TTA AAT GAT TTA TGT ATC AAA GTT Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 465 470 475 480	1440

AAT AAT TGG GAC TTG TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 485 490 495	1488
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 500 505 510	1536
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 515 520 525	1584
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 530 535 540	1632
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 545 550 555 560	1680
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 565 570 575	1728
CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe His Gly Lys Ser Arg Ile Ala Leu Thr 580 585 590	1776
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 595 600 605	1824
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 610 615 620	1872
ATG TTT TTA GCC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 625 630 635 640	1920
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 645 650 655	1968
ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 660 665 670	2016
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 675 680 685	2064
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 690 695 700	2112
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 705 710 715 720	2160
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 725 730 735	2208
GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 740 745 750	2256

- 39 -

AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 755 760 765	2304
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 770 775 780	2352
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 785 790 795 800	2400
AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 805 810 815	2448
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 820 825 830	2496
GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 835 840 845	2544
AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 850 855 860	2592
AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 865 870 875 880	2640
GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 885 890 895	2685

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 895 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu
 1 5 10 15

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn
 20 25 30

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro
 35 40 45

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys
 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly
 65 70 75 80

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr
 85 90 95

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys
 100 105 110

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg
 115 120 125
 Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser
 130 135 140
 Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val
 145 150 155 160
 Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile
 165 170 175
 Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly
 180 185 190
 His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr
 195 200 205
 Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu
 210 215 220
 Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro
 225 230 235 240
 Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr
 245 250 255
 Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala
 260 265 270
 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr
 275 280 285
 Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu
 290 295 300
 Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu
 305 310 315 320
 Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met
 325 330 335
 Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly
 340 345 350
 Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu
 355 360 365
 Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu
 370 375 380
 Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn
 385 390 395 400
 Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg
 405 410 415
 Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn
 420 425 430
 Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe
 435 440 445
 Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser
 450 455 460

- 41 -

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val
 465 470 475 480
 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 485 490 495
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 500 505 510
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 515 520 525
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 530 535 540
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 545 550 555 560
 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 565 570 575
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 580 585 590
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 595 600 605
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 610 615 620
 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 625 630 635 640
 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 645 650 655
 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 660 665 670
 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 675 680 685
 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 690 695 700
 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 705 710 715 720
 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 725 730 735
 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 740 745 750
 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 755 760 765
 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 770 775 780
 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 785 790 795 800
 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 805 810 815

- 42 -

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 820 825 830

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 835 840 845

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 850 855 860

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 865 870 875 880

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys
 885 890 895

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2622 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2622

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGA TCC ATG GAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val 1 5 10 15	48
AAC GGT GTT GAC ATT GCC TAC ATC AAA ATT CCA AAG TAC GGC CAG ATG Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30	96
CAG CCG GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro 35 40 45	144
GAA CGC GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro 50 55 60	192
CCG GAA GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu 65 70 75 80	240
AGC ACA GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu 85 90 95	288
TTC GAG CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser 100 105 110	336
ATC GTC CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu 115 120 125	384
TTG AAG GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly 130 135 140	432

- 43 -

AGC TAC AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG · Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala 145 150 155 160	480
GAC ATT ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn 165 170 175	528
CTG ACG CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro 180 185 190	576
GAC TTC ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 195 200 205	624
CTG TTG GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala 210 215 220	672
CAC GAG CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn 225 230 235 240	720
CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 245 250 255	768
GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 260 265 270	816
GCG AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 275 280 285	864
TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 290 295 300	912
ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 305 310 315 320	960
GAG AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 325 330 335	1008
AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 345 350	1056
GAG GAT AAT TTT GTT AAG TTT AAA GTA CTT AAC AGA AAA ACA TAT Glu Asp Asn Phe Val Lys Phe Lys Val Leu Asn Arg Lys Thr Tyr 355 360 365	1104
TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 370 375 380	1152
AAT TAC ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 385 390 395 400	1200
GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 405 410 415	1248

AAA CTA AAA AAT TTT ACT GGA TTG TTI GAA TTT TAT AAG TTG CTA TGT Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys 420 425 430	1296
GTA AGA GGG AIA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr 435 440 445	1344
AAT AAG GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu 450 455 460	1392
TTT TTT ACT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 465 470 475 480	1440
GAA GAA ATT ACA TCT GAT ACT AAT AIA GAA GCA GCA GAA GAA AAT ATT Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 485 490 495	1488
AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn 500 505 510	1536
GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly 515 520 525	1584
CAA TTA GAA CCT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys 530 535 540	1632
TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu 545 550 555 560	1680
TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu 555 570 575	1728
GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr 580 585 590	1776
GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp 595 600 605	1824
GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser 610 615 620	1872
ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly 625 630 635 640	1920
CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT GAT TTT GTA GGT Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly 645 650 655	1968
GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu 660 665 670	2016
ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 675 680 685	2064

- 45 -

AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg 690 695 700	2112
AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu 705 710 715 720	2160
GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu 725 730 735	2208
GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln 740 745 750	2256
TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile 755 760 765	2304
GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile 770 775 780	2352
AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn 785 790 795 800	2400
TCT ATG ATC CCT TAT GGT GTT AAA CCG TTA GAA GAT TTT GAT GCT AGT Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser 805 810 815	2448
CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu 820 825 830	2496
ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser 835 840 845	2544
ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu 850 855 860	2592
TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Leu Ser Thr Phe Thr Glu Tyr Ile Lys *	2622
865 870	

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 874 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val	
1 5 10 15	
Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met	
20 25 30	
Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro	
35 40 45	

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro
 50 55 60
 Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu
 65 70 75 80
 Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu
 85 90 95
 Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser
 100 105 110
 Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu
 115 120 125
 Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly
 130 135 140
 Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala
 145 150 155 160
 Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn
 165 170 175
 Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro
 180 185 190
 Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro
 195 200 205
 Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala
 210 215 220
 His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn
 225 230 235 240
 Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser
 245 250 255
 Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp
 260 265 270
 Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr
 275 280 285
 Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser
 290 295 300
 Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys
 305 310 315 320
 Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp
 325 330 335
 Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr
 340 345 350
 Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr
 355 360 365
 Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val
 370 375 380
 Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala
 385 390 395 400

- 47 -

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr
 405 410 415
 Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys
 420 425 430
 Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr
 435 440 445
 Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu
 450 455 460
 Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly
 465 470 475 480
 Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile
 485 490 495
 Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn
 500 505 510
 Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly
 515 520 525
 Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys
 530 535 540
 Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu
 545 550 555 560
 Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu
 565 570 575
 Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr
 580 585 590
 Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp
 595 600 605
 Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser
 610 615 620
 Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly
 625 630 635 640
 Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly
 645 650 655
 Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu
 660 665 670
 Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala
 675 680 685
 Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg
 690 695 700
 Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu
 705 710 715 720
 Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu
 725 730 735
 Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln
 740 745 750

- 48 -

Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile
 755 760 765

Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile
 770 775 780

Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn
 785 790 795 800

Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser
 805 810 815

Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu
 820 825 830

Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser
 835 840 845

Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu
 850 855 860

Leu Ser Thr Phe Thr Glu Tyr Ile Lys
 865 870

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2613 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2613

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATG CCA TTT GTT AAT AAA CAA TTT AAT TAT AAA GAT CCT GTA AAT GGT	48
Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
1 5 10 15	
GTT GAT ATT GCT TAT ATA AAA ATT CCA AAT GCA GGA CAA ATG CAA CCA	96
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
20 25 30	
GTA AAA GCT TTT AAA ATT CAT AAT AAA ATA TGG GTT ATT CCA GAA AGA	144
Val Lys Ala Phe Lys Ile His Asn Lys Ile Tyr Val Ile Pro Glu Arg	
35 40 45	
GAT ACA TTT ACA AAT CCT GAA GAA GGA GAT TTA AAT CCA CCA CCA GAA	192
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu	
50 55 60	
GCA AAA CAA GTT CCA GTT TCA TAT TAT GAT TCA ACA TAT TTA AGT ACA	240
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	
65 70 75 80	
GAT AAT GAA AAA GAT AAT TAT TTA AAG GGA GTT ACA AAA TTA TTT GAG	288
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	
85 90 95	

- 49 -

AGA ATT TAT TCA ACT GAT CTT GGA AGA ATG TTG TTA ACA TCA ATA GTA Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
AGG GGA ATA CCA TTT TGG GGT GGA AGT ACA ATA GAT ACA GAA TTA AAA Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
GTT ATT GAT ACT AAT TGT ATT AAT GTG ATA CAA CCA GAT GGT AGT TAT Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
AGA TCA GAA GAA CTT AAT CTA GTA ATA ATA GGA CCC TCA GCT GAT ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
ATA CAG TTT GAA TGT AAA AGC TTT GGA CAT GAA GTT TTG AAT CTT ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
CGA AAT GGT TAT GGC TCT ACT CAA TAC ATT AGA TTT AGC CCA GAT TTT Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
ACA TTT GGT TTT GAG GAG TCA CTT GAA GTT GAT ACA AAT CCT CTT TTA Thr Phe Gly Phe Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAA TTT GCT ACA GAT CCA GCA GTA ACA TTA GCA CAT GAA Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTT ATA CAT GCT GGA CAT AGA TTA TAT GGA ATA GCA ATT AAT CCA AAT Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
AGG GTT TTT AAA GTA AAT ACT AAT GCC TAT TAT GAA ATG AGT GGG TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTT GAG GAA CTT AGA ACA TTT GGG GGA CAT GAT GCA AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTT ATA GAT AGT TTA CAG GAA AAC GAA TTT CGT CTA TAT TAT TAT AAT Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATA GCA AGT ACA CTT AAT AAA GCT AAA TCA ATA GTA Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACT ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104

- 50 -

TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT CGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 455 460	1392
AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480	1440
ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA AAT ATT AGT TTA Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	1488
GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510	1536
GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560	1680
CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640	1920

TTA AAT ATA GGT AAI ATG TTA TAT AAA GAT GAT TTT GTA GGT GCT TTA Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655	1968
ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720	2160
GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735	2208
GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750	2256
CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT GAT GAT Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765	2304
TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780	2352
AAT AAA TTT TTG AAT CAA TGC TCT GTA TAT TTA ATG AAT TCT ATG Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800	2400
ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT CTT AAA Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815	2448
GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA ATT GGT Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830	2496
CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT ACA GAT Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845	2544
ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA TTA TCT Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860	2592
ACA TTT ACT GAA TAT ATT AAG Thr Phe Thr Glu Tyr Ile Lys 865 870	2613

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 871 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

- 52 -

(ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335

- 53 -

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
 450 455 460
 Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480
 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525
 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540
 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560
 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575
 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590
 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605
 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620
 Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640
 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655
 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670
 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685

- 54 -

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750

Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
 805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
 850 855 860

Thr Phe Thr Glu Tyr Ile Lys
 865 870

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2628 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2628

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT	48
Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
1 5 10 15	
GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG	96
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
20 25 30	
GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC	144
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
35 40 45	
GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA	192
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	
50 55 60	

- 55 -

GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	268
CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008

AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
AGC GCT GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp 450 455 460	1392
GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480	1440
AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA GAA Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 495	1488
AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 505 510	1536
GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile 515 520 525	1584
ATA GGC CAA TTA GAA CCT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly 530 535 540	1632
AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 550 555 560	1680
CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 570 575	1728
AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser 580 585 590	1776
GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 600 605	1824

GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 615 620	1872
GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640	1920
ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650 655	1968
GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile 660 665 670	2016
CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr 675 680 685	2064
ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser 690 695 700	2112
AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn 705 710 715 720	2160
TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met 725 730 735	2208
AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn 740 745 750	2256
TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe 755 760 765	2304
AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala 770 775 780	2352
ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu 785 790 795 800	2400
ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp 805 810 815	2448
GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820 825 830	2496
ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr 835 840 845	2544
CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln 850 855 860	2592
AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875	2628

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 876 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met	Gln	Phe	Val	Asn	Lys	Gln	Phe	Asn	Tyr	Lys	Asp	Pro	Val	Asn	Gly	
1															15	
Val	Asp	Ile	Ala	Tyr	Ile	Lys	Ile	Pro	Asn	Ala	Gly	Gln	Met	Gln	Pro	
													20	25	30	
Val	Lys	Ala	Phe	Lys	Ile	His	Asn	Lys	Ile	Trp	Val	Ile	Pro	Glu	Arg	
													35	40	45	
Asp	Thr	Phe	Thr	Asn	Pro	Glu	Glu	Gly	Asp	Leu	Asn	Pro	Pro	Pro	Glu	
												50	55	60		
Ala	Lys	Gln	Val	Pro	Val	Ser	Tyr	Tyr	Asp	Ser	Thr	Tyr	Leu	Ser	Thr	
													65	70	75	80
Asp	Asn	Glu	Lys	Asp	Asn	Tyr	Leu	Lys	Gly	Val	Thr	Lys	Leu	Phe	Glu	
												85	90	95		
Arg	Ile	Tyr	Ser	Thr	Asp	Leu	Gly	Arg	Met	Leu	Leu	Thr	Ser	Ile	Val	
												100	105	110		
Arg	Gly	Ile	Pro	Phe	Trp	Gly	Gly	Ser	Thr	Ile	Asp	Thr	Glu	Leu	Lys	
												115	120	125		
Val	Ile	Asp	Thr	Asn	Cys	Ile	Asn	Val	Ile	Gln	Pro	Asp	Gly	Ser	Tyr	
												130	135	140		
Arg	Ser	Glu	Glu	Leu	Asn	Leu	Val	Ile	Ile	Gly	Pro	Ser	Ala	Asp	Ile	
												145	150	155	160	
Ile	Gln	Phe	Glu	Cys	Lys	Ser	Phe	Gly	His	Glu	Val	Leu	Asn	Leu	Thr	
												165	170	175		
Arg	Asn	Gly	Tyr	Gly	Ser	Thr	Gln	Tyr	Ile	Arg	Phe	Ser	Pro	Asp	Phe	
												180	185	190		
Thr	Phe	Gly	Phe	Glu	Glu	Ser	Leu	Glu	Val	Asp	Thr	Asn	Pro	Leu	Leu	
												195	200	205		
Gly	Ala	Gly	Lys	Phe	Ala	Thr	Asp	Pro	Ala	Val	Thr	Leu	Ala	His	Glu	
												210	215	220		
Leu	Ile	His	Ala	Gly	His	Arg	Leu	Tyr	Gly	Ile	Ala	Ile	Asn	Pro	Asn	
												225	230	235	240	
Arg	Val	Phe	Lys	Val	Asn	Thr	Asn	Ala	Tyr	Tyr	Glu	Met	Ser	Gly	Leu	
												245	250	255		
Glu	Val	Ser	Phe	Glu	Glu	Leu	Arg	Thr	Phe	Gly	Gly	His	Asp	Ala	Lys	
												260	265	270		
Phe	Ile	Asp	Ser	Leu	Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	Tyr	Tyr	Asn	
												275	280	285		
Lys	Phe	Lys	Asp	Ile	Ala	Ser	Thr	Leu	Asn	Lys	Ala	Lys	Ser	Ile	Val	
												290	295	300		

- 59 -

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Ph Ser Val Asp Lys Leu
325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435 440 445

Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp
450 455 460

Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn
465 470 475 480

Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu
485 490 495

Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe
500 505 510

Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile
515 520 525

Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly
530 535 540

Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala
545 550 555 560

Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val
565 570 575

Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser
580 585 590

Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu
595 600 605

Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu
610 615 620

Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr
625 630 635 640

Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe
645 650 655

- 60 -

Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile
 660 665 670

Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr
 675 680 685

Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser
 690 695 700

Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn
 705 710 715 720

Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met
 725 730 735

Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn
 740 745 750

Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe
 755 760 765

Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala
 770 775 780

Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu
 785 790 795 800

Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp
 805 810 815

Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly
 820 825 830

Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr
 835 840 845

Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln
 850 855 860

Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys
 865 870 875

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2637 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE_TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2637

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT
 Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 1 5 10 15

48

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG
 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
 20 25 30

96

GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
GAT ACA TTT ACG AAC CCG GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60	192
GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912

- 62 -

GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680
CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728

- 63 -

AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA .TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920
ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544

- 64 -

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA
Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
850 855 860

GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA
Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys *
865 870 875

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 879 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr

www.Ban-Sky.Tom/Sky-Song-The-Sky-Tom-Ho-Ban-Sky-Song/Ban-Sky-Song

Sky-Ne-Sky-Luc-Rho-Ne-Thru-Nam-Pres-Ne-Vol-Thru-Lan-Ne-Wic-Sky

210 215 220

225 **230** **235** **240**

- 65 -

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270
 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285
 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300
 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460
 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605

- 66 -

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu			
610	615	620	
Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile			
625	630	635	640
Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys			
645	650	655	
Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu			
660	665	670	
Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu			
675	680	685	
Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn			
690	695	700	
Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile			
705	710	715	720
Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg			
725	730	735	
Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala			
740	745	750	
Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn			
755	760	765	
Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile			
770	775	780	
Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val			
785	790	795	800
Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu			
805	810	815	
Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp			
820	825	830	
Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val			
835	840	845	
Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val			
850	855	860	
Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys *			
865	870	875	

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2862 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2862

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

- 67 -

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15	48
GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30	96
GTC AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60	192
GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816

- 68 -

TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632

- 69 -

CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680
CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920
ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
GTA ACA AAT TCG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448

- 70 -

GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544
AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860	2592
GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880	2640
CCT GGA CCG GAG ACG CTC TGC GGG GCT GAG CTG GTG GAT GCT CTT CAG Pro Gly Pro Glu Thr Leu Cys Gly Ala Glu Leu Val Asp Ala Leu Gln 885 890 895	2688
TTC GTG TGT GGA GAC AGG GGC TTT TAT TTC AAC AAG CCC ACA GGG TAT Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe Asn Lys Pro Thr Gly Tyr 900 905 910	2736
GGC TCC AGC AGT CGG AGG GCG CCT CAG ACA GGT ATC GTG GAT GAG TGC Gly Ser Ser Arg Arg Ala Pro Gln Thr Gly Ile Val Asp Glu Cys 915 920 925	2784
TGC TTC CGG AGC TGT GAT CTA AGG AGG CTG GAG ATG TAT TGC GCA CCC Cys Phe Arg Ser Cys Asp Leu Arg Arg Leu Glu Met Tyr Cys Ala Pro 930 935 940	2832
CTC AAG CCT GCC AAG TCA GCT GAA GCT TAG Leu Lys Pro Ala Lys Ser Ala Glu Ala * 945 950	2862

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 954 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95
Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

- 71 -

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445

Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460

- 72 -

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605
 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620
 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640
 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655
 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 660 665 670
 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 675 680 685
 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 690 695 700
 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn
 755 760 765
 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 775 780
 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 785 790 795 800
 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 805 810 815

- 73 -

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 820 825 830
 Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 835 840 845
 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860
 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg
 865 870 875 880
 Pro Gly Pro Glu Thr Leu Cys Gly Ala Glu Leu Val Asp Ala Leu Gln
 885 890 895
 Phe Val Cys Gly Asp Arg Gly Phe Tyr Phe Asn Lys Pro Thr Gly Tyr
 900 905 910
 Gly Ser Ser Ser Arg Arg Ala Pro Gln Thr Gly Ile Val Asp Glu Cys
 915 920 925
 Cys Phe Arg Ser Cys Asp Leu Arg Arg Leu Glu Met Tyr Cys Ala Pro
 930 935 940
 Leu Lys Pro Ala Lys Ser Ala Glu Ala
 945 950

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2724 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2724

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT	48
Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	
1 5 10 15	
GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG	96
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro	
20 25 30	
GTC AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC	144
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	
35 40 45	
GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA	192
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	
50 55 60	
GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA	240
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	
65 70 75 80	
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG	288
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	
85 90 95	

- 74 -

CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104

- 75 -

TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
AAT AAT TGG GAC TTG TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680
CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920

- 76 -

ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
AAT AAA CCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544
AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860	2592
GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCT AGG Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880	2640
CCT CAA TCT AAA GTT AAA AGA CAA ATA TTT TCA GGC TAT CAA TCT GAT Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 885 890 895	2688
ATT GAT ACA CAT AAT AGA ATT AAG GAT GAA TTA TGA Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu 900 905	2724

- 77 -

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 908 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300

- 78 -

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320
 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335
 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350
 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365
 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380
 Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400
 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415
 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445
 Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460
 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480
 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495
 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510
 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525
 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540
 Pro Asn Gly Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560
 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575
 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590
 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605
 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620
 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640
 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655

- 79 -

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 660 665 670
 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 675 680 685
 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 690 695 700
 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 755 760 765
 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 775 780
 Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 785 790 795 800
 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 805 810 815
 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 820 825 830
 Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 835 840 845
 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860
 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg
 865 870 875 880
 Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp
 885 890 895
 Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu *
 900 905

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3042 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..3042

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT
 Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
 1 5 10 15

- 80 -

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30	96
GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45	144
GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60	192
GCA AAG CAG GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80	240
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95	288
CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110	336
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125	384
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140	432
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160	480
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175	528
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190	576
ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864

- 81 -

AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460	1392
AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480	1440
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1488
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510	1536
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	1584
AGT GAC ATT ATA CGC CAA TTA GAA CCT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540	1632
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560	1680

- 82 -

CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575	1728
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620	1872
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920
ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 645 650 655	1968
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670	2016
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685	2064
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700	2112
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720	2160
GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765	2304
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	2400
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496

- 83 -

AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544
AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860	2592
GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TCA GGC Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 865 870 875 880	2640
CTG AAT TCC CCG GGT GCA GCT CAT TAT GCG CAA CAC GAT GAA GCC GTA Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895	2688
GAC AAC AAA TTC AAC AAA GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Ley 900 905 910	2736
CAT TTA CCT AAC TTA AAC GAA GAA CAA CGA AAC GCC TTC ATC CAA AGT His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925	2784
TTA AAA GAT GAC CCA AGC CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940	2832
AAG CTA AAT GAT GCT CAG GCG CCG AAA GTA GAC AAC AAA TTC AAC AAA Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960	2880
GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975	2928
GAA GAA CAA CGA AAC GCC TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990	2976
CAA AGC GCT AAC CTT TTA GCA GAA GCT AAA AAG CTA AAT GAT GCT CAG Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005	3024
GCG CCG AAA GTA GAC TAG Ala Pro Lys Val Asp * 1010	3042

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1014 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

- 84 -

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
 385 390 395 400

- 85 -

Phe Asn Gly Gln Asn Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435 440 445

Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val
 450 455 460

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn
 465 470 475 480

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala
 485 490 495

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
 500 505 510

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser
 515 520 525

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe
 530 535 540

Pro Asn Gly Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr
 545 550 555 560

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr
 565 570 575

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe
 580 585 590

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala
 595 600 605

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu
 610 615 620

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile
 625 630 635 640

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu
 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750

- 86 -

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 775 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
 805 810 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
 850 855 860

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly
 865 870 875 880

Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val
 885 890 895

Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu
 900 905 910

His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser
 915 920 925

Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys
 930 935 940

Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys
 945 950 955 960

Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn
 965 970 975

Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser
 980 985 990

Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln
 995 1000 1005

Ala Pro Lys Val Asp
 1010

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3509 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..3509

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

- 87 -

ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	48
AAT AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60	192
ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80	240
ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160	480
TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	528
ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190	576
ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220	672
GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240	720
GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816

- 88 -

GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320	960
AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400	1200
TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445	1344
GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460	1392
GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480	1440
TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488
TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510	1536
GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525	1584
AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535 540	1632

- 89 -

ACA TTT CCT CTA GAT ATA AGA GAT ATA AGT TTA ACA TCT TCA TTT GAT Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 555 560	1680
GAT GCA TTA TTA TTT TCT AAC AAA GTT TAT TCA TTT TTT TCT ATG GAT Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575	1728
TAT ATT AAA ACT GCT AAT AAA GTG GTA GAA GCA GGA TTA TTT GCA GGT Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590	1776
TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AGC Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605	1824
AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT ATA Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 615 620	1872
GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GAA Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 630 635 640	1920
AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CCA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655	1968
GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATT Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670	2016
GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AAA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685	2064
AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TGG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700	2112
CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TAT Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720	2160
AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TAC Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735	2208
AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GAT Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750	2256
TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT ATA Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765	2304
GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATG Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 775 780	2352
AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAT Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800	2400
ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815	2448

- 90 -

TTG ATT GGA AGT GCA GAA TAT GAA AAA TCA AAA GTA AAT AAA TAC TTG Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830	2496
AAA ACC ATT ATG CCG TTT GAT CTT TCA ATA TAT ACC AAT GAT ACA ATA Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845	2544
CTA ATA GAA ATG TTT AAT AAA TAT AAT AGC GAA ATT TTA AAT AAT ATT Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855 860	2592
ATC TTA AAT TTA AGA TAT AAG GAT AAT AAT TTA ATA GAT TTA TCA GGA Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 865 870 875 880	2640
TAT GGG GCA AAG GTA GAG GTA TAT GAT GGA GTC GAG CTT AAT GAT AAA Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895	2688
AAT CAA TTT AAA TTA ACT AGT TCA GCA AAT AGT AAG ATT AGA GTG ACT Asn Gln Phe Lys Leu Thr Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910	2736
CAA AAT CAG AAT ATC ATA TTT AAT AGT GTG TTC CTT GAT TTT AGC GTT Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925	2784
AGC TTT TGG ATA AGA ATA CCT AAA TAT AAG AAT GAT GGT ATA CAA AAT Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940	2832
TAT ATT CAT AAT GAA TAT ACA ATA ATT AAT TGT ATG AAA AAT AAT TCG Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960	2880
GGC TGG AAA ATA TCT ATT AGG GGT AAT AGG ATA ATA TGG ACT TTA ATT Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975	2928
GAT ATA AAT GGA AAA ACC AAA TCG GTA TTT TTT GAA TAT AAC ATA AGA Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990	2976
GAA GAT ATA TCA GAG TAT ATA AAT AGA TGG TTT TTT GTA ACT ATT ACT Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005	3024
AAT AAT TTG AAT AAC GCT AAA ATT TAT ATT AAT GGT AAG CTA GAA TCA Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020	3072
AAT ACA GAT ATT AAA GAT ATA AGA GAA GTT ATT GCT AAT GGT GAA ATA Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040	3120
ATA TTT AAA TTA GAT GGT GAT ATA GAT AGA ACA CAA TTT ATT TGG ATG Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055	3168
AAA TAT TTC AGT ATT TTT AAT ACG GAA TTA AGT CAA TCA AAT ATT GAA Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070	3216
GAA AGA TAT AAA ATT CAA TCA TAT AGC GAA TAT TTA AAA GAT TTT TGG Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085	3264

- 91 -

GGA AAT CCT TTA ATG TAC AAT AAA GAA TAT TAT ATG TTT AAT GCG GGG Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100	3312
AAT AAA AAT TCA TAT ATT AAA CTA AAG AAA GAT TCA CCT GTA GGT GAA Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120	3360
ATT TTA ACA CGT AGC AAA TAT AAT CAA AAT TCT AAA TAT ATA AAT TAT Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135	3408
AGA GAT TTA TAT ATT GGA GAA AAA TTT ATT ATA AGA AGA AAG TCA AAT Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150	3456
TCT CAA TCT ATA AAT GAT GAT ATA GTT AGA AAA GAA GAT TAT ATA TAT Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165	3504
CFA GA Leu	3509

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1169 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15
Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30
Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45
Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60
Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80
Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95
Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110
Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125
Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140
Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160
Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

- 92 -

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
 180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu
 195 200 205

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
 210 215 220

Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
 225 230 235 240

Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe
 245 250 255

Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe
 260 265 270

Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile
 275 280 285

Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn
 290 295 300

Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr
 305 310 315 320

Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly
 325 330 335

Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu
 340 345 350

Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys
 355 360 365

Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
 370 375 380

Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
 385 390 395 400

Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile
 405 410 415

Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
 420 425 430

Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
 435 440 445

Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser
 450 455 460

Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn
 465 470 475 480

Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
 485 490 495

Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
 500 505 510

Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
 515 520 525

- 93 -

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540 |
 Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp
 545 550 555 560
 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp
 565 570 575
 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly
 580 585 590
 Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser
 595 600 605
 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile
 610 615 620
 Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu
 625 630 635 640
 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro
 645 650 655
 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile
 660 665 670
 Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys
 675 680 685
 Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp
 690 695 700
 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr
 705 710 715 720
 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr
 725 730 735
 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp
 740 745 750
 Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile
 755 760 765
 Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met
 770 775 780
 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn
 785 790 795 800
 Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr
 805 810 815
 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu
 820 825 830
 Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile
 835 840 845
 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile
 850 855 860
 Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly
 865 870 875 880

- 94 -

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys
 885 890 895
 Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr
 900 905 910
 Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val
 915 920 925
 Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn
 930 935 940
 Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser
 945 950 955 960
 Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile
 965 970 975
 Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg
 980 985 990
 Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr
 995 1000 1005
 Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser
 1010 1015 1020
 Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile
 1025 1030 1035 1040
 Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met
 1045 1050 1055
 Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu
 1060 1065 1070
 Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp
 1075 1080 1085
 Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly
 1090 1095 1100
 Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu
 1105 1110 1115 1120
 Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr
 1125 1130 1135
 Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn
 1140 1145 1150
 Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr
 1155 1160 1165
 Leu

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 95 -

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	48
AAT AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60	192
ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80	240
ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160	480
TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	528
ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190	576
ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GGT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220	672
GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240	720

- 96 -

GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
TTT ATG CAA TCT ACA GAT GCT ATA CAG GCA GAA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816
GGA CGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320	960
AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400	1200
TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445	1344
GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460	1392
GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480	1440
TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488
TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510	1536

- 97 -

GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525	1584
AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535 540	1632
ACA TTT CCT CTA GAT ATA AGA GAT ATA AGT TTA ACA TCT TCA TTT GAT Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 555 560	1680
GAT GCA TTA TTA TTT TCT AAC AAA GTT TAT TCA TTT TTT TCT ATG GAT Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575	1728
TAT ATT AAA ACT GCT AAT AAA GTG GTA GAA GCA GGA TTA TTT GCA GGT Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590	1776
TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AGC Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605	1824
AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT ATA Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 615 620	1872
GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GAA Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 630 635 640	1920
AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CCA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655	1968
GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATT Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670	2016
GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AAA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685	2064
AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TGG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700	2112
CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TAT Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720	2160
AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TAC Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735	2208
AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GAT Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750	2256
TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT ATA Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755 760 765	2304
GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATG Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 775 780	2352

- 98 -

AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAT Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800	2400
ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815	2448
TTG ATT GGA AGT GCA GAA TAT GAA AAA TCA AAA GTA AAT AAA TAC TTG Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830	2496
AAA ACC ATT ATG CCG TTT GAT CTT TCA ATA TAT ACC AAT GAT ACA ATA Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845	2544
CTA ATA GAA ATG TTT AAT AAA TAT AAT AGC Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser 850 855	2574

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 858 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15
Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30
Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45
Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60
Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80
Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95
Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110
Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125
Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140
Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160
Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175
Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190

- 99 -

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu
 195 200 205 |

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
 210 215 220

Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
 225 230 235 240

Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe
 245 250 255

Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe
 260 265 270

Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile
 275 280 285

Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn
 290 295 300

Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr
 305 310 315 320

Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly
 325 330 335

Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu
 340 345 350

Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys
 355 360 365

Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
 370 375 380

Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
 385 390 395 400

Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile
 405 410 415

Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
 420 425 430

Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
 435 440 445

Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser
 450 455 460

Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn
 465 470 475 480

Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
 485 490 495

Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
 500 505 510

Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
 515 520 525

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540

- 100 -

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp
 545 550 555 560
 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp
 565 570 575
 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly
 580 585 590
 Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser
 595 600 605
 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile
 610 615 620
 Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu
 625 630 635 640
 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro
 645 650 655
 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile
 660 665 670
 Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys
 675 680 685
 Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp
 690 695 700
 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr
 705 710 715 720
 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr
 725 730 735
 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp
 740 745 750
 Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile
 755 760 765
 Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met
 770 775 780
 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn
 785 790 795 800
 Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr
 805 810 815
 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu
 820 825 830
 Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile
 835 840 845
 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser
 850 855

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1644 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

- 101 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ATG CCA GTT ACA ATA AAT AAT TTT AAT TAT AAT GAT CCT ATT GAT AAT Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15	48
AAT AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGG AGA Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30	96
TAT TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45	144
AGA TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60	192
ATT TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80	240
ACT AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA TTT Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95	288
AAT AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG ATG ATT Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110	336
ATA AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA GAG Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125	384
TTT AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT AAT Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140	432
CCA GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATA ATA Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160	480
TTT GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT ATA GGT Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175	528
ATA CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA ATG CAA Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190	576
ATG AAG TTT TGC CCA GAA TAT GTA AGC GTA TTT AAT AAT GTT CAA GAA Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205	624
AAC AAA GGC GCA AGT ATA TTT AAT AGA CGT GGA TAT TTT TCA GAT CCA Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220	672

- 102 -

GCC TTG ATA TTA ATG CAT GAA CTT ATA CAT GTT TTA CAT GGA TTA TAT Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 225 230 235 240	720
GGC ATT AAA GTA GAT GAT TTA CCA ATT GTA CCA AAT GAA AAA AAA TTT Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255	768
TTT ATG CAA TCT ACA GAT CCT ATA CAG GCA GAA CTA TAT ACA TTT Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270	816
GGA GGA CAA GAT CCC AGC ATC ATA ACT CCT TCT ACG GAT AAA AGT ATC Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285	864
TAT GAT AAA GTT TTG CAA AAT TTT AGA GGG ATA GTT GAT AGA CTT AAC Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300	912
AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATT AAT ATA TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 305 310 315 320	960
AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 330 335	1008
AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350	1056
ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365	1104
ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 375 380	1152
AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400	1200
TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415	1248
AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430	1296
AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 440 445	1344
GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460	1392
GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480	1440
TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495	1488

- 103 -

TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510	1536
GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525	1584
AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535 540	1632
ACA TTT CCT CTA Thr Phe Pro Leu 545	1644

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 548 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15
Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30
Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45
Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 60
Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80
Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95
Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110
Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125
Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140
Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160
Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175
Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190
Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205

- 104 -

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro
 210 215 220
 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr
 225 230 235 240
 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe
 245 250 255
 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe
 260 265 270
 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile
 275 280 285
 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn
 290 295 300
 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr
 305 310 315 320
 Lys Asn Lys Phe Lys Asp Tyr Lys Phe Val Glu Asp Ser Glu Gly
 325 330 335
 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu
 340 345 350
 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys
 355 360 365
 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys
 370 375 380
 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile
 385 390 395 400
 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile
 405 410 415
 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr
 420 425 430
 Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp
 435 440 445
 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser
 450 455 460
 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn
 465 470 475 480
 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp
 485 490 495
 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr
 500 505 510
 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys
 515 520 525
 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln
 530 535 540
 Thr Phe Pro Leu
 545

- 105 -

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2616 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly	48
1 5 10 15	
GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Pro Asn Ala Gly Gln Met Gln Pro	96
20 25 30	
GTC AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg	144
35 40 45	
GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu	192
50 55 60	
GCA AAG CAC GTG CCA GTT TCA TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr	240
65 70 75 80	
GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu	288
85 90 95	
CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val	336
100 105 110	
CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys	384
115 120 125	
GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr	432
130 135 140	
AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile	480
145 150 155 160	
ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr	528
165 170 175	
CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	576
180 185 190	

- 108 -

ACG TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG CTG TTG Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205	624
GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA CAC GAG Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220	672
CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC CCG AAC Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240	720
CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT GGT TTA Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255	768
GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GCG AAG Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270	816
TTC ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC TAC AAC Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285	864
AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC ATT GTG Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GAG AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT AAA TTA Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA GAG GAT Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350	1056
AAT TTT GTT AAG TTT AAA GTA CTT AAC AGA AAA ACA TAT TTG AAT Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400	1200
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CTA Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445	1344
GCA TTA AAT GAT TTA TGT ATC AAA GTT AAT AAT TGG GAC TTG TTT TTT Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 455 460	1392

- 107 -

AGT CCT TCA GAA GAT AAT TTT ACT AAT GAT CTA AAT AAA GGA GAA GAA Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480	1440
ATT ACA TCT GAT ACT AAT ATA GAA GCA GCA GAA AAT ATT AGT TTA Ile Thr Ser Asp Thr Asn Ile Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495	1488
GAT TTA ATA CAA CAA TAT TAT TTA ACC TTT AAT TTT GAT AAT GAA CCT Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510	1536
GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525	1584
GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540	1632
TIA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560	1680
CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575	1728
TIA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG Leu Asn Pro Ser Arg Val Tyr Thr Phe Ser Ser Asp Tyr Val Lys 580 585 590	1776
AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605	1824
CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620	1872
GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640	1920
TIA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT TTT GTA GGT GCT TTA Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655	1968
ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670	2016
ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685	2064
GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700	2112
AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720	2160
GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735	2208

- 108 -

GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750	2256
CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT AAT ATT GAT GAT Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765	2304
TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780	2352
AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA ATG AAT TCT ATG Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800	2400
ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT GCT AGT CTT AAA Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815	2448
GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA ACT TTA ATT GGT Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830	2496
CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT AAT ACA CTT AGT ACA GAT Gln Val Asp Arg Leu Lys Asp Val Asn Asn Thr Leu Ser Thr Asp 835 840 845	2544
ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA AGA TTA TTA TCT Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860	2592
ACA TTT ACT GAA TAT ATT AAG TAA Thr Phe Thr Glu Tyr Ile Lys * 865 870	2616

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 872 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 5 10 15
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95
Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

- 109 -

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125 |
Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140
Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145 150 155 160
Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
165 170 175
Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
180 185 190
Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
195 200 205
Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
210 215 220
Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn
225 230 235 240
Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu
245 250 255
Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
260 265 270
Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn
275 280 285
Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val
290 295 300
Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys
305 310 315 320
Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu
325 330 335
Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp
340 345 350
Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn
355 360 365
Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr
370 375 380
Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn
385 390 395 400
Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu
405 410 415
Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
420 425 430
Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435 440 445
Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe
450 455 460

- 110 -

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu
 465 470 475 480

Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495

Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro.
 500 505 510

Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
 805 810 815

- 111 -

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
 850 855 860

Thr Phe Thr Glu Tyr Ile Lys *
 865 870

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2574 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:..

ATGCCGGTAA	CCATCAACAA	CTTCAACTAC	AACGACCCGA	TCGACAACAA	CAACATCATC	60
ATGATGGAAC	CGCCGTTCCGC	ACGTGGTACC	GGTCGTTACT	ACAAGGCTTT	CAAGATCACC	120
GACCGTATCT	GGATCATCCC	GGAACGTTAC	ACCTTCGGTT	ACAAACCTGA	GGACTTCAAC	180
AAGAGTAGCG	GGATTTCAA	TCGTGACGTC	TGGGAGTACT	ATGATCCAGA	TTATCTGAAT	240
ACCAACGATA	AGAAGAACAT	ATTCTTCAG	ACTATGATCA	AGTTATTTAA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATAACC	TTATCTTGGAA	360
GATAGACGTG	TTCCACTCGA	AGAGTTAAC	ACAAACATTG	CTAGTGTAAAC	TGTTAATAAA	420
TTAACATCGA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAAGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCTG	GGCCAGTTTT	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTT	CGGGGGTATA	ATGCAAATGA	AGTTTGCCC	AGAATAATGTA	600
ACCGTATTTA	ATAATGTTCA	AGAAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTCAGATC	CAGCCTTGAT	ATTAATGCAT	GAACTTATAC	ATGTTTACA	TGGATTATAT	720
GGCATTAAAG	TAGATGATT	ACCAATTGTA	CCAAATGAAA	AAAAATTTTT	TATGCAATCT	780
ACAGATGCTA	TACACCCAGA	AGAACTATAT	ACATTGGAG	GACAAGATCC	CAGCATCATA	840
ACTCCTTCTA	CGGATAAAAG	TATCTATGAT	AAAGTTTGC	AAAATTTAG	AGGGATAGTT	900
GATAGACTTA	ACAAGGTTTT	AGTTTGCATA	TCAGATCCTA	ACATTAATAT	TAATATATAT	960
AAAAATAAAAT	TTAAAGATAA	ATATAAATT	GTTGAAGATT	CTGAGGGAAA	ATATAGTATA	1020
GATGTAGAAA	GTTTGATAA	ATTATATAAA	AGCTTAATGT	TTGGTTTAC	AGAAAACAAAT	1080
ATAGCAGAAA	ATTATAAAAT	AAAAACTAGA	GCTTCTTATT	TTAGTGATTC	CTTACCAACCA	1140
GTAAAAATAA	AAAATTTATT	AGATAATGAA	ATCTACTA	TAGAGGAAGG	TTTTAATATA	1200

- 112 -

TCTGATAAAG ATATGGAAAA AGAATATAGA GGTCAGAATA AAGCTATAAA TAAACAAAGCT 1260
 TATGAAGAAA TTAGCAAGGA GCATTGGCT GTATATAAGA TACAAATGTG TAAAAGTGT 1320
 AAAGCTCCAG GAATATGTAT TGATGTTGAT AATGAAGATT TGTTCTTAT AGCTGATAAA 1380
 AATAGTTTT CAGATGATT ATCTAAAAC GAAAGAATAG AATATAATAC ACAGAGTAAT 1440
 TATATAGAAA ATGACTTCCC TATAAATGAA TTAATTTAG ATACTGATT AATAAGTAAA 1500
 ATAGAATTAC CAAGTGAAAA TACAGAATCA CTACTGATT TTAATGTAGA TGTTCCAGTA 1560
 TATGAAAAAC AACCGCTAT AAAAAAAATT TTTACAGATG AAAATACCAT CTTTCAATAT 1620
 TTATACCTTC AGACATTTC TCTAGATATA AGAGATATAA GTTTAACATC TTCAATTGAT 1680
 GATGCATTAT TATTTCTAA CAAAGTTTAT TCATTTTTT CTATGGATTA TATTAAAATC 1740
 GCTAATAAAG TGGTAGAAGC AGGATTATT GCAGGTTGGG TGAAACAGAT AGTAAATGAT 1800
 TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT 1860
 GTTCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTGAA 1920
 AATGCTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTAATA 1980
 CCTGTAGTTG GAGCTTTT ATTAGAATCA TATATTGACA ATAAAATAA AATTATTAAA 2040
 ACAATAGATA ATGCTTAAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA 2100
 GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTATA CAATAAAAGA GGGAAATGTAT 2160
 AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA 2220
 TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTAA ATGATATAAA TTCTAAACTT 2280
 AATGAGGGTA TTAACCAAGC TATAGATAAT ATAAATAATT TTATAAATGG ATGTTCTGTA 2340
 TCATATTAA TGAAAAAAAT GATTCCATTA GCTGTAGAAA AATTACTAGA CTTTGATAAT 2400
 ACTCTCAAAA AAAATTGTT AAATTATATA GATGAAAATA AATTATATT GATTGGAAGT 2460
 GCAGAATATG AAAATCAAAGTAAATAGA TACTTGAAA CCATTATGCC GTTTGATCTT 2520
 TCAATATATA CCAATGATAC AATACTAATA GAAATGTTA ATAAATATAA TAGC 2574

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2574 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ATGCCAGTTA CAATAAATAA TTTTAATTAT AATGATCCTA TTGATAATAA TAATATTATT 60
 ATGATGGAGC CTCCATTGTC GAGAGGTACG GGGAGATATT ATAAAGCTTT TAAAATCACA 120
 GATCGTATTG GGATAATACC GGAAAGATAT ACTTTGGAT ATAAACCTGA GGATTTAAT 180
 AAAAGTTCCG GTATTTTAA TAGAGATGTT TGTGAATATT ATGATCCAGA TTACTTAAAT 240

- 113 -

ACTAATGATA AAAAGAATAT/ATTTTACAA ACAATGATCA AGTTATTTAA TAGAATCAA	300
TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	360
GATAGACGTG TTCCACTCGA AGAGTTAAC ACAAAACATTG CTAGTGTAAAC TGTTAATAAAA	420
TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTAA TTTCGCAAATTTAATAATA	480
TTTGGACCTG GGCCAGTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	540
TTTGCATCAA GGGAAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTGCCC AGAATATGTAA	600
AGCGTATTAA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACgtGGATAT	660
TTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACgtCCTCCA CGGTCTCTAC	720
GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAAATTCTT CATGCAGAGC	780
ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	840
ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCTCC AGAACTTCCG TGGTATCGTC	900
GACCGTCTCA ACAAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	960
AAAAACAAAT TCAAAGACAA ATACAAATTG GTCGAAGACA CGGAAGGTAA ATACAGCATC	1020
GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1080
ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1140
GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTCAACATC	1200
ACCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACAA AAGCAATCAA CAAACAAAGCT	1260
TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAAA TCCAGATGTG CAAAAGCGTC	1320
AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1380
AACAGCTTCA GCGACGACCT CAGCAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1440
TACATCGAAA ACCACTTCCC GATCAACGAA CTCATCTCG ACACCGACCT CATCAGCAAA	1500
ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
TACGAAAAC AGCCGGCAAT CAAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTGAT	1680
GATGCATTAT TATTTCTAA CAAAGTTTAT TCATTTTTT CTATGGATTA TATTTAAACT	1740
GCTAATAAAG TGGTAGAAGC AGGATTATT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
TTTGTAAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGAGATAT ATCTCTAATT	1860
GTTCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAAATTTGAA	1920
AATGCTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTAAATA	1980
CCTGTAGTTG GAGCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTATA CAATAAAAGA GGGAAATGTAT	2160
AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTAAATGATATAAA TTCTAAACTT	2280

- 114 -

AATGAGGGTA TTAACCAAGC TATAGATAAT ATAATAATT TTATAAATGG ATGTTCTGTA 2340
TCATATTTAA TGAAAAAAAAT GATTC CATT A GCTGTAGAAA AATTACTAGA CTTTGATAAT 2400
ACTCTCAAAA AAAATTGTT AAATTATATA GATGAAAATA AATTATATT GATTGGAAGT 2460
GCAGAATATG AAAATCAA AGTAAATAAA TACTTGAAA CCATTATGCC GTTGATCTT 2520
TCAATATATA CCAATGATAC AATACTAATA GAAATGTTA ATAAATATAA TAGC 2574

- 115.-

CLAIMS

1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain H_N portion.
7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain H_N portion.
8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain β of Staphylococcal protein A.
12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated H_N of a botulinum toxin heavy chain.
15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 117 -

17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H_c of a botulinum toxin heavy chain.

26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
27. A polypeptide according to Claim 26 lacking an intact portion designated H_c of a botulinum toxin heavy chain.
28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
33. A fusion protein according to Claim 32 wherein said second polypeptide is

- 119 -

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.

35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell-surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.

36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.

37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.

38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.

39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.

40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.

41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

- 120 -

40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H_N domain.

43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.

44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H_N domain.

45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.

46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.

47. A DNA according to any of claims 40-46.

48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.

49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.

50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

- 121 -

according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.

1/11

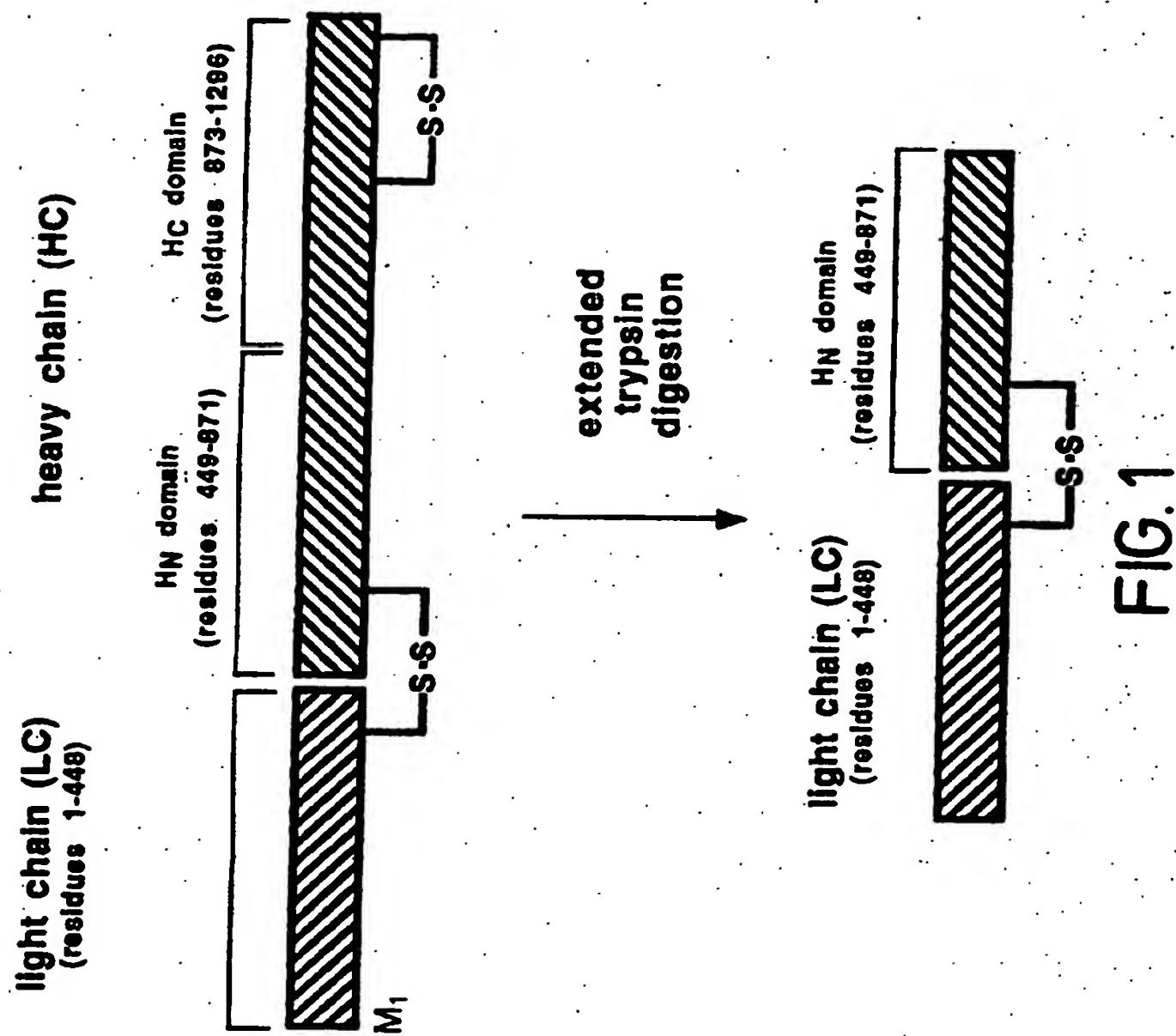


FIG. 1

2 / 11

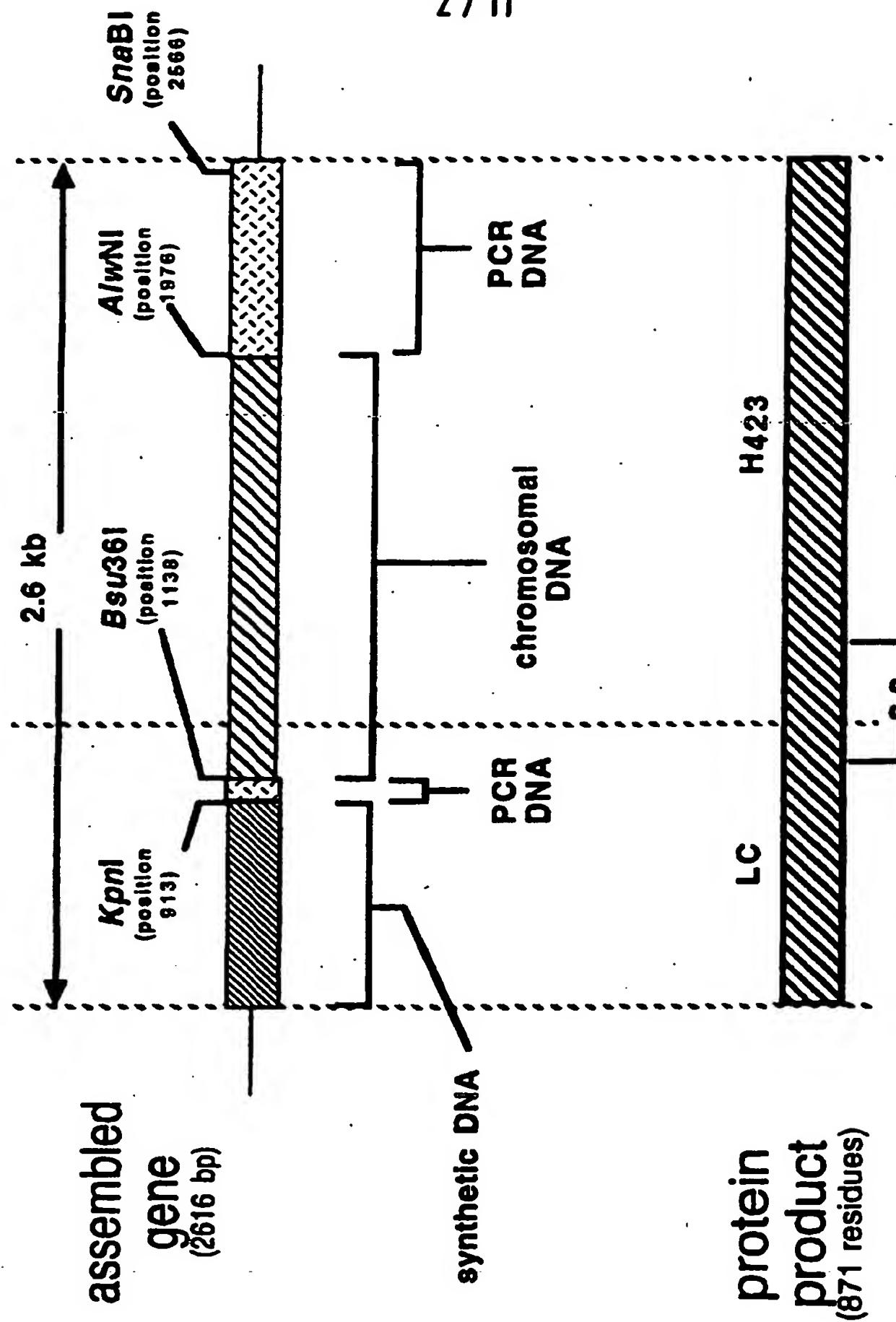


FIG. 2

3 / 11

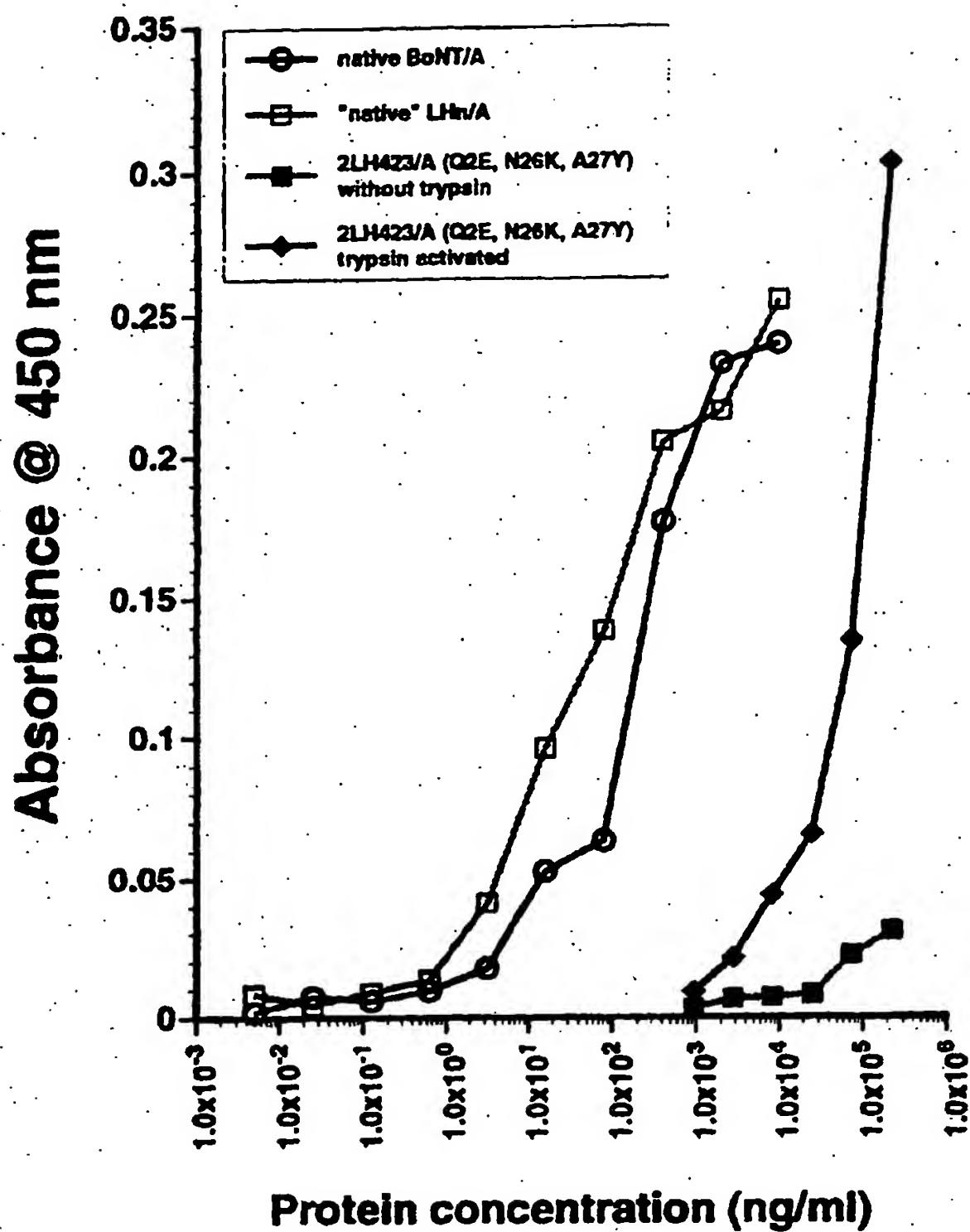


FIG. 3

LH_{423/A}
M Q F V N K Q F N Y K D P V N G V D I A Y I K I P N A G Q M Q P V (Seq I.D. 2)

1

[G S P G I H M T S T R L Q K L L E F E L P

**23LH_{423/A}
(Q₂E, N₂₆K, A₂₇Y)**

G T N E F V N K Q F N Y K D P V N G V D I A Y I K I P K Y G Q M Q P V (Seq I.D. 4)
24

**2LH_{423/A}
(Q₂E, N₂₆K, A₂₇Y)**

G S M E F V N K Q F N Y K D P V N G V D I A Y I K I P K Y G Q M Q P V (Seq I.D. 6) ↳ 11

**Native BoNT/A,
C. botulinum 2169
Thompson et al. 1990**

**Native BoNT/A,
C. botulinum 62A
Binz et al. 1990**

[M Q F V N K Q F N Y K D P V N G V D I A Y I K I P N A G Q M Q P V
1



= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.

FIG. 4

5 / 11

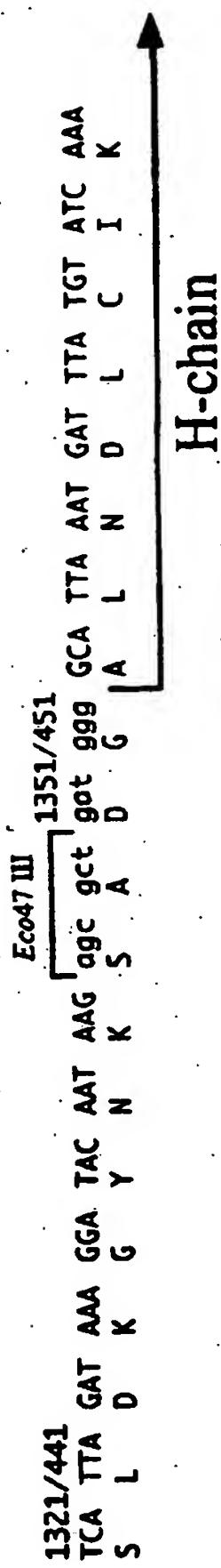


FIG. 5

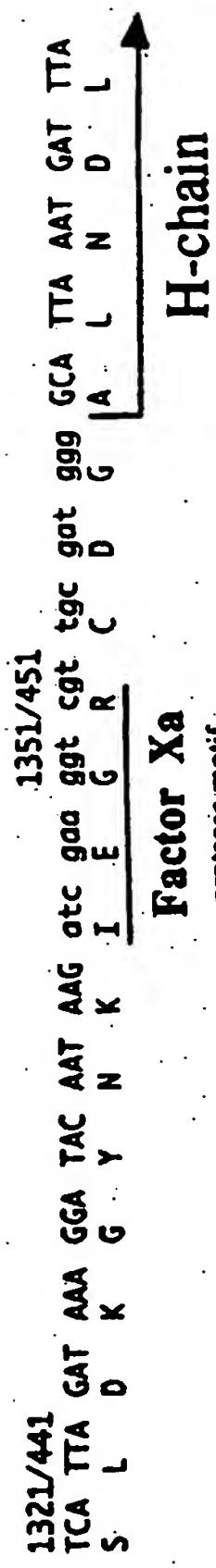


FIG. 6

6/11

IGF-1 →

2587/863	TAC	GTA	GAT	AAT	CAA	AGA	TTA	TTA	TCT	ACA	TTT	ACT	GAA	TAT	ATT	AAG	TCT	AGG	CCT	GGA	
Y	V	D	N	Q	R	L	S	T	F	T	E	Y	I	K	S	R	P	G			
2647/883	CCC	GAG	ACG	CTC	TGC	GGC	GCT	GAC	CTG	GAT	GCT	CTT	CAG	TTC	GTC	TGT	GGA	GAC	AGG		
P	E	T	L	C	G	A	E	L	V	D	A	L	Q	F	V	C	G	D	R		
2707/903	GGC	TCC	TTC	TAT	TTC	AAC	AAG	CCC	ACA	GGG	TAT	GGC	TCC	AGC	AGT	CGG	AGG	GCG	CCT	CAG	ACA
G	F	Y	F	N	K	P	T	G	Y	G	S	S	R	R	A	P	Q	Q	T		
2767/923	GGT	ATC	GTC	GAT	GAG	TGC	TGC	TTC	CGG	AGC	TGT	GAT	CTA	AGG	AGG	CTG	GAG	ATG	TAT	TGC	
G	I	V	D	E	C	C	F	R	S	C	D	L	R	L	E	M	Y	C			
2827/943	GCA	CCC	CTC	AAG	CCT	GGC	AAG	TCA	GCT	GAA	GCT	TAG									
A	P	L	K	P	A	K	S	A	E	A	stop										

FIG. 7

CtxA14 →

2587/863	TAC	GTA	GAT	AAT	CAA	AGA	TTA	TTA	TCT	ACA	TTT	ACT	GAA	TAT	ATT	AAG	TCT	AGG	CCT	GAA	
Y	V	D	N	Q	R	L	S	T	F	T	E	Y	I	K	S	R	P	Q			
2647/883	TCT	AAA	GTT	AAA	AGA	CAA	ATA	TTT	TCA	GGC	TAT	CAA	TCT	GAT	ATT	GAT	ACA	CAT	AAT	AGA	
S	K	V	K	R	Q	I	F	S	G	Y	Q	S	D	I	D	T	H	N	R		
2707/903	ATT	AAG	GAT	GAA	TTA	TGA															
I	K	D	E	L	stop																

FIG. 8

7/11

2587/863	TAC GAT AAT CAA AGA TTA TCT ACA	TTT ACT GAA TAT ATT AAG TCA	GGC CTG AAT	
Y V D N Q R L	S T F T E Y I K S G L	N		
2647/883	TCC CCC GGT GCA GCT CAT TAT GCG CAA CAC GAT GAA	GCC GTC GAC AAC AAA	TTC AAC AAA	
S P G A A H	A Q H D E A V D N K	F N K		
<u>2707/903</u>	GAA CAA CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC GAA GAA CAA CGA			
E Q Q N A F	E I L H L P N L	N E E Q R		
2767/923	AAC GCC TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC CAA GCT AAC CTT TTA GCA GAA			
N A F I Q S	K D D S A	N L L A E		
2827/943	GCT AAA AAG CTA AAT GAT GCT CAG GCG CCG AAA GTC GAC AAC AAA TTC AAC AAC CAA			
A K K L N D	A Q A P K V D N K F N K	E E Q		
2887/963	CAA AAC GCG TTC TAT GAG ATC TTA CAT TTA CCT AAC TTA AAC GAA GAA CAA CGA AAC			
Q N A F	I L H L P N L	N E E Q R		
2947/983	TTC ATC CAA AGT TTA AAA GAT GAC CCA AGC GCT AAC CTT TTA GCA GAA GCT AAA			
F I Q S	L K D D P S Q S A	N L L A E A K		
3007/1003	AAG CTA AAT GAT GCT CAG GCG CCC AAA GTA GAC TAG			3037/1013
K L N D	A Q A P K V D			*

FIG. 9

8/11
LH_{423/A}

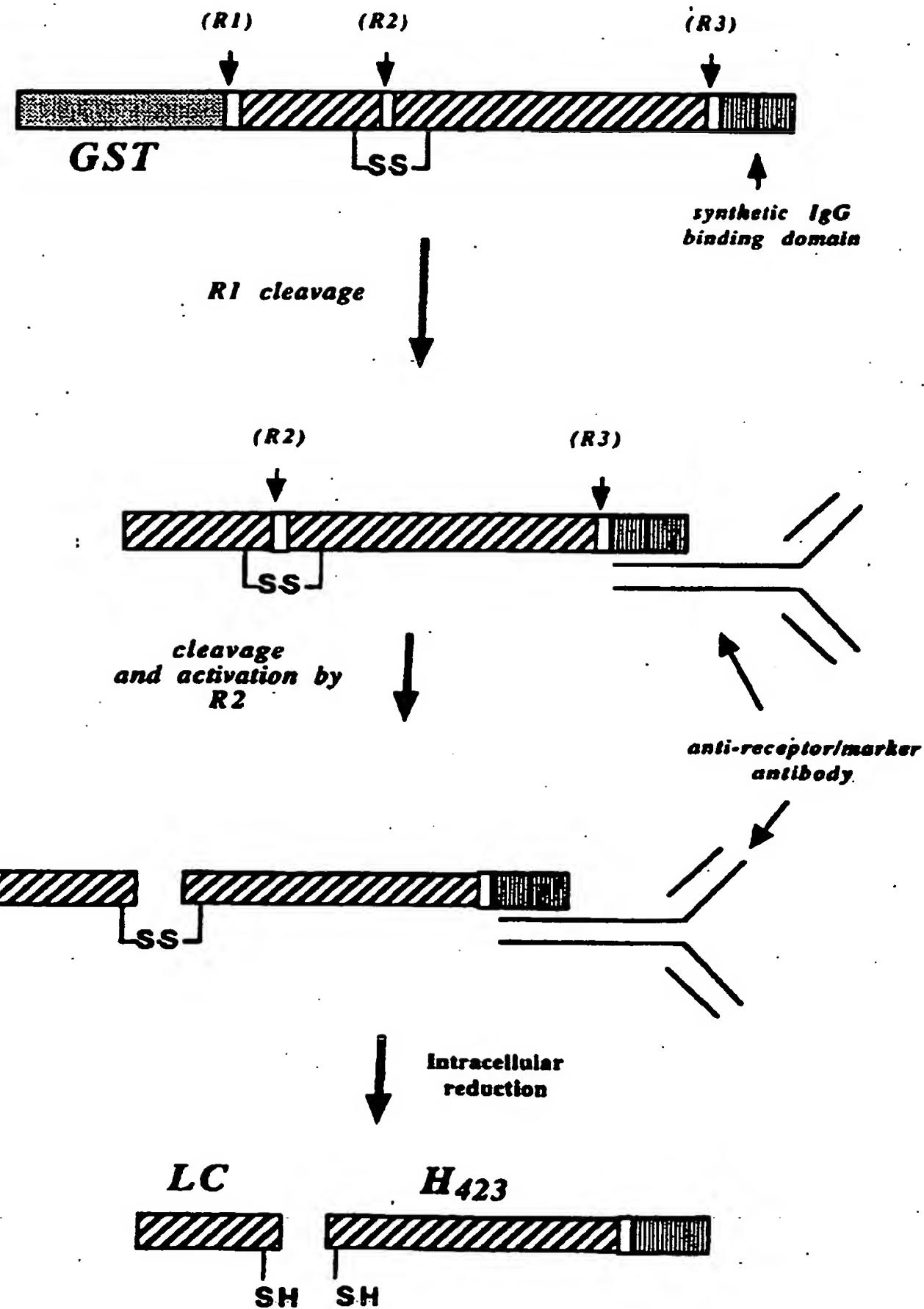


FIG. 10

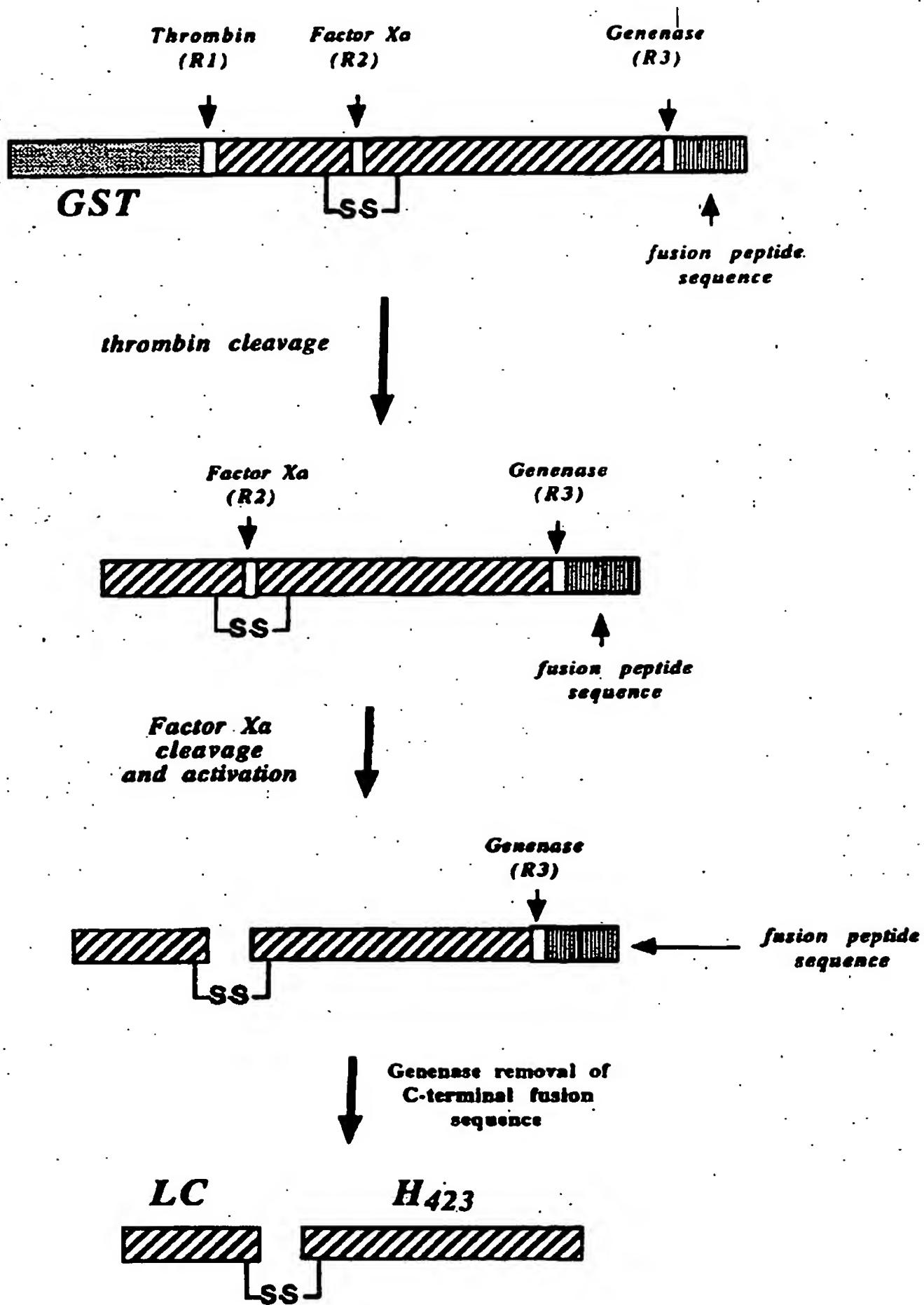
LH₄₂₃/A^{9/11}

FIG. 11

10 / 11

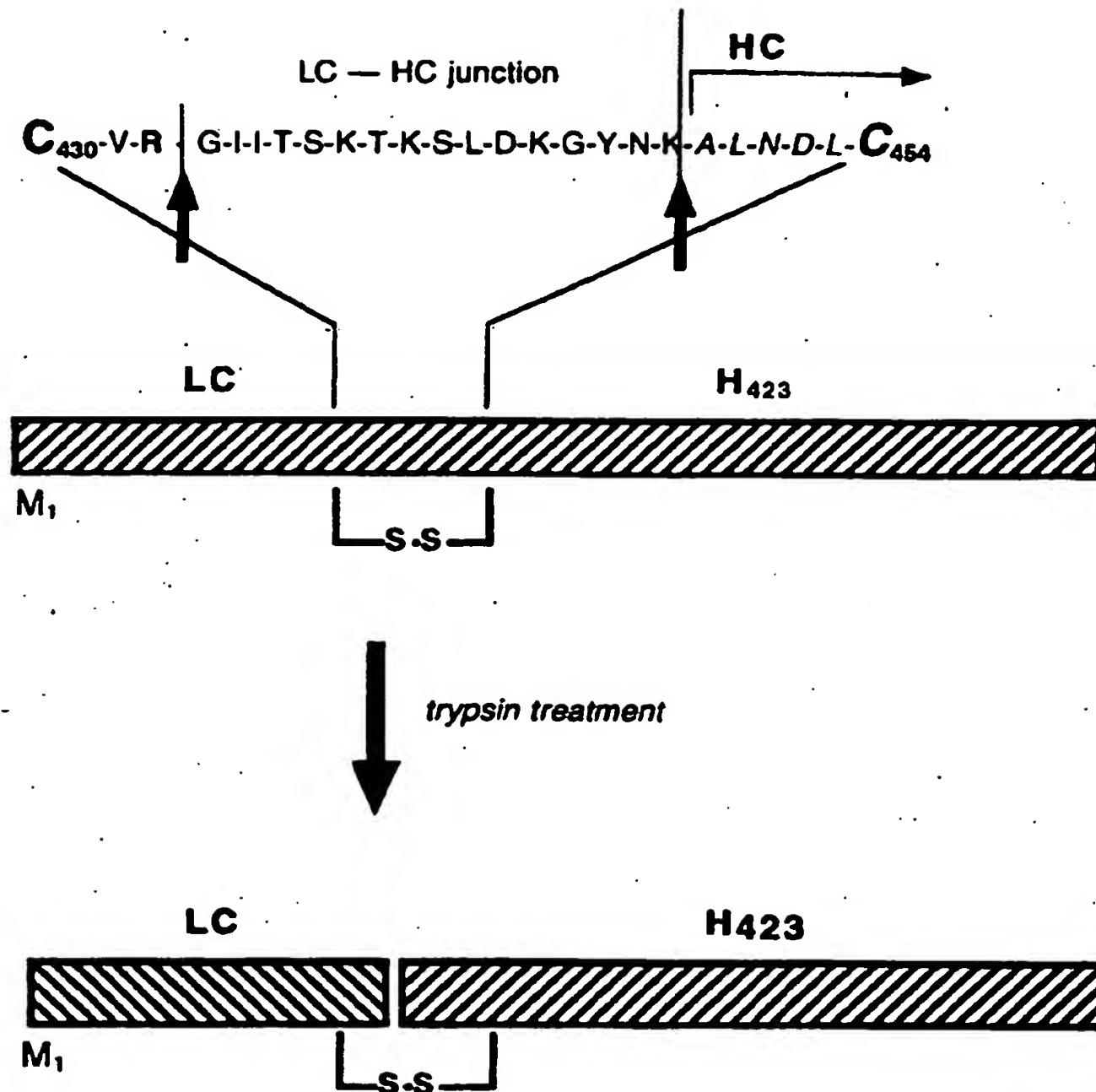
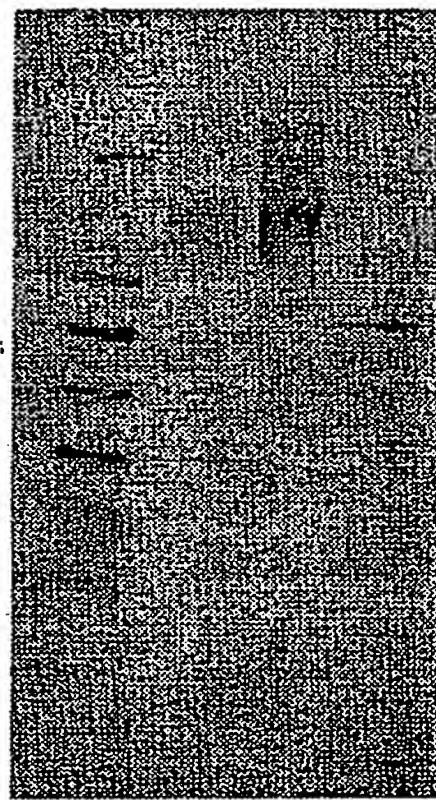


FIG. 12

11 / 11

Panel A.

1 2 3 4



Panel B.

1 2 3 4

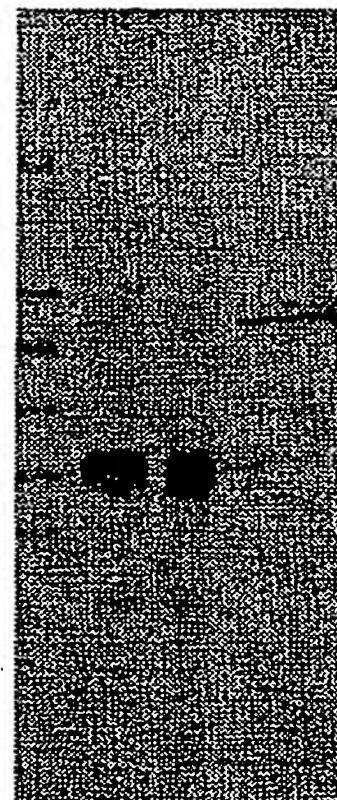


FIG. 13

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 97/02273

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/31 C12N1/21 C12P21/02 C07K14/33 A61K38/16
A61K39/08

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12P A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 12892 A (OPHIDIAN PHARM INC ;WILLIAMS JAMES A (US); PADHYE NISHA V (US); KJ) 2 May 1996 see the whole document ---	1-52
X	KURAZONO H ET AL: "Minimal essential *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II --- -/-	1-52

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

Z document member of the same patent family

Date of the actual completion of the international search

9 December 1997

Date of mailing of the international search report

30.01.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patenttaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hillenbrand, G

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 97/02273

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document	1
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document	1,26,35

INTERNATIONAL SEARCH REPORT

Int'l. Appl. No.
PCT/GB 97/02273

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9612892 A	02-05-96	US 5601823 A	11-02-97
		US 5196193 A	23-03-93
		AU 3968395 A	15-05-96
		EP 0796326 A	24-09-97
		FI 971732 A	23-06-97
		NO 971868 A	24-06-97
		PL 328214 A	15-09-97
		ZA 9508990 A	15-05-96
		AU 6653894 A	04-07-94
		CA 2150935 A	23-06-94
		EP 0671902 A	20-09-95
		WO 9413264 A	23-06-94
		US 5466672 A	14-11-95
		US 5599539 A	04-02-97
		AU 638786 B	08-07-93
		AU 6895191 A	31-05-91
		EP 0498854 A	19-08-92
		WO 9106306 A	16-05-91
		US 5443976 A	22-08-95
		US 5340923 A	23-08-94